

PERBANDINGAN KINERJA MODEL REGRESI LOGISTIK, SUPPORT VECTOR MACHINE, DAN NAÏVE BAYES DALAM MEMPREDIKSI KEPUASAN MAHASISWA TERHADAP PROSES PEMBELAJARAN

Trigarcia Maleachi Randa^{1,*}, Loria Amisah Lubis²

^{1,2}Program Studi Matematika, Fakultas Matematika Dan Ilmu Pengetahuan Alam, Universitas Papua

*Email Korespondensi: t.randa@unipa.ac.id

ABSTRACT

This study aims to analyze and compare the performance of three classification methods Logistic Regression, Support Vector Machine (SVM), and Naïve Bayes in predicting student satisfaction with the learning process during the 2025/2026 odd semester at the University of Papua. The research data were collected through a survey consisting of eight categorical predictor variables related to aspects of the learning process, while the response variable comprised two categories: satisfied and not satisfied. The analysis was conducted by splitting the dataset into training and testing sets with a proportion of 80:20. The results indicate that Logistic Regression and SVM achieved the highest accuracy, each at 95.24%, whereas the Naïve Bayes method produced an accuracy of 87.88%, despite applying a 5-fold cross-validation procedure and Laplace smoothing to determine the optimal parameters. These findings suggest that Logistic Regression and SVM are the most effective methods for predicting student satisfaction in this dataset, while Naïve Bayes remains an efficient alternative for simple and fast modeling. The results of this study are expected to support decision-making in evaluating and improving the quality of learning in higher education institutions.

Keywords: Logistic Regression, Support Vector Machine, Naïve Bayes, Student Satisfaction.

ABSTRACT

Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja tiga metode klasifikasi Regresi Logistik, Support Vector Machine (SVM), dan Naïve Bayes dalam memprediksi kepuasan mahasiswa terhadap proses pembelajaran pada semester gasal 2025/2026 di Universitas Papua. Data penelitian diperoleh melalui survei yang berisi delapan variabel prediktor kategorik yang berkaitan dengan aspek proses pembelajaran, sementara variabel respon terdiri atas dua kategori, yaitu puas dan tidak puas. Analisis dilakukan dengan membagi data menjadi training dan testing dengan proporsi 80:20. Hasil penelitian menunjukkan bahwa metode Regresi Logistik dan SVM memberikan akurasi tertinggi, masing-masing sebesar 95.24%, sedangkan metode Naïve Bayes menghasilkan akurasi sebesar 87.88% meskipun telah dilakukan penentuan parameter terbaik menggunakan validasi silang *5-fold* dan penerapan *Laplace smoothing*. Temuan ini menunjukkan bahwa Regresi Logistik dan SVM merupakan metode yang paling efektif untuk memprediksi kepuasan mahasiswa pada dataset ini, sementara Naïve Bayes tetap menjadi alternatif yang efisien untuk pemodelan yang sederhana dan cepat. Hasil penelitian ini diharapkan dapat mendukung pengambilan keputusan dalam evaluasi dan peningkatan kualitas pembelajaran di perguruan tinggi.

Keywords: Regresi Logistik, Support Vector Machine, Naïve Bayes, Kepuasan Mahasiswa

ARTICLE INFO

Submission received: 22 November 2025

Accepted: 30 December 2025

Revised: 15 December 2025

Published: 31 December 2025

Available on: <https://doi.org/10.32493/sm.v7i3.54904>

StatMat: Jurnal Statistika dan Matematika is licenced under a Creative Commons Attribution-ShareAlike 4.0 International License.



1. PENDAHULUAN

Kepuasan mahasiswa terhadap proses pembelajaran merupakan salah satu indikator penting dalam mengevaluasi kualitas pendidikan tinggi. Tingkat kepuasan yang baik tidak hanya mencerminkan efektivitas proses belajar-mengajar, tetapi juga menjadi dasar bagi perguruan tinggi untuk melakukan perbaikan berkelanjutan dalam penyediaan layanan akademik. Faktor-faktor seperti kualitas dosen, fasilitas pendukung, dan aspek pembelajaran lainnya turut mempengaruhi persepsi mahasiswa terhadap pengalaman belajar yang mereka peroleh (Roni & Atmaja, 2020). Dalam konteks pendidikan moderen, kemampuan untuk memprediksi kepuasan mahasiswa menjadi semakin penting agar institusi dapat mengidentifikasi area kelemahan secara lebih cepat dan tepat, sehingga peningkatan mutu pembelajaran dapat dilakukan secara terarah.

Seiring berkembangnya teknologi dan ketersediaan data, berbagai pendekatan analitik dan model prediktif mulai diterapkan dalam evaluasi pendidikan. Metode-metode machine learning menjadi pilihan populer karena kemampuannya dalam mengolah data beragam, menangkap pola yang kompleks, dan menghasilkan prediksi dengan akurasi tinggi (Roihan et al., 2020). Di antara metode yang umum digunakan, regresi logistik, Support Vector Machine (SVM), dan Naïve Bayes merupakan tiga algoritma klasifikasi yang banyak diterapkan dalam penelitian bidang pendidikan maupun bidang lainnya.

Regresi logistik merupakan metode statistik yang digunakan untuk memodelkan hubungan antara variabel respon biner dan variabel prediktor, serta mampu memberikan estimasi probabilitas terhadap suatu kejadian. Metode ini populer karena interpretasinya yang jelas dan kemampuannya untuk mengevaluasi pengaruh masing-masing prediktor terhadap hasil (Amalia et al., 2025). Dalam konteks kepuasan mahasiswa, regresi logistik dapat digunakan untuk mengidentifikasi faktor-faktor pembelajaran yang secara signifikan mempengaruhi kemungkinan mahasiswa merasa puas atau tidak puas.

Support Vector Machine (SVM) merupakan algoritma klasifikasi yang bekerja dengan mencari hyperplane terbaik untuk memisahkan dua kelas dengan margin maksimum. SVM dikenal memiliki kinerja yang baik terutama pada data berdimensi tinggi dan pola yang tidak linier melalui penggunaan *kernel trick* (Fahrezi et al., 2024). Dengan demikian, SVM menjadi salah satu metode yang kuat dalam memodelkan dinamika kepuasan mahasiswa yang mungkin dipengaruhi oleh kombinasi variabel yang kompleks. Sementara itu, Naïve Bayes merupakan metode klasifikasi probabilistik berbasis Teorema Bayes dengan asumsi independensi antar fitur. Meskipun sederhana dan memiliki asumsi yang kuat, Naïve Bayes terbukti efisien dan sering memberikan hasil yang kompetitif, terutama pada data kategorik maupun data berukuran besar (Aristawidya, 2024). Kesederhanaan perhitungan dan kecepatan pemrosesan menjadi keunggulan utama metode ini dalam analisis prediktif.

Beberapa penelitian sebelumnya telah membahas penerapan metode-metode tersebut dalam berbagai konteks. Misalnya, Sari et al. (2024) menunjukkan bahwa regresi logistik memiliki performa klasifikasi yang lebih baik dibandingkan naïve bayes dalam prediksi penerima bantuan Program Keluarga Harapan (PKH). Penelitian lain oleh Cahyani et al. (2022) menekankan bahwa proses normalisasi data mampu meningkatkan performa regresi logistik dalam memprediksi risiko diabetes. Namun, penelitian yang membandingkan ketiga metode regresi logistik, SVM, dan Naïve Bayes secara simultan dalam konteks prediksi kepuasan mahasiswa masih terbatas. Oleh karena itu, penelitian ini bertujuan untuk membandingkan kinerja ketiga metode tersebut dalam memprediksi kepuasan mahasiswa terhadap proses pembelajaran pada semester gasal 2025/2026 di Universitas Papua. Penelitian ini akan mengevaluasi kedua model berdasarkan beberapa metrik kinerja seperti akurasi, sensitivitas, spesifisitas, dan presisi, baik pada data training maupun testing. Hasil

penelitian ini diharapkan dapat memberikan gambaran yang jelas mengenai metode yang paling efektif digunakan untuk memprediksi kepuasan mahasiswa, serta memberikan kontribusi terhadap pengembangan model evaluasi pembelajaran berbasis data di lingkungan pendidikan tinggi.

2. METODOLOGI

2.1 Dataset dan Variabel

Kajian ini menggunakan desain penelitian kuantitatif dengan metode klasifikasi berbasis *machine learning* untuk memprediksi kepuasan mahasiswa terhadap proses pembelajaran. Fokus utamanya adalah membandingkan kinerja tiga algoritma pembelajaran terawasi, yaitu Regresi Logistik, Naïve Bayes, dan Support Vector Machine (SVM), dalam mengklasifikasikan tingkat kepuasan mahasiswa berdasarkan data survei yang tersedia. Data sekunder yang digunakan berasal dari hasil survei kepuasan mahasiswa terhadap proses pembelajaran di Universitas Papua pada Semester Gasal Tahun Akademik 2025/2026. Informasi tersebut diperoleh melalui kuesioner daring yang disebarakan kepada mahasiswa dari berbagai program studi, dengan total responden sebanyak 108 orang.

Variabel respon dalam penelitian ini adalah kepuasan mahasiswa terhadap proses pembelajaran, yang dikategorikan menjadi dua kelas, yaitu “puas” dan “tidak puas”. Sementara itu, variabel prediktor yang digunakan mencakup beberapa aspek proses pembelajaran, seperti kualitas dosen, metode pengajaran, fasilitas pendukung, interaksi dengan dosen, beban tugas, kesesuaian waktu pembelajaran, kondisi lingkungan belajar, dan motivasi mahasiswa. Variabel-variabel tersebut diukur dalam bentuk data kategorik (nominal) dan dijelaskan secara rinci pada Tabel 1.

Tabel 1. Variabel Penelitian

Variabel	Kategori
Kepuasan Mahasiswa (Y)	0: Tidak Puas 1: Puas
Kualitas Dosen (X_1)	1: Kurang 2: Cukup 3: Baik
Metode Mengajar (X_2)	1: Ceramah 2: Diskusi 3: Praktikum 4: Multimedia
Fasilitas Pendukung (X_3)	1: Memadai 2: Tidak Memadai
Interaksi dengan Dosen (X_4)	1: Jarang 2: Kadang 3: Sering
Beban Tugas (X_5)	1: Ringan 2: Sedang 3: Berat
Waktu Pembelajaran (X_6)	1: Sesuai 0: Tidak Sesuai
Motivasi Mahasiswa (X_7)	1: Rendah 2: Sedang 3: Tinggi

2.1 Regresi Logistik

Regresi Logistik merupakan model statistik yang digunakan untuk memperkirakan probabilitas suatu outcome biner berdasarkan satu atau lebih variabel prediktor (Hosmer et al., 2013). Berbeda dengan regresi linier yang memprediksi nilai kontinu, regresi logistik memprediksi kemungkinan terjadinya suatu peristiwa menggunakan fungsi logistik:

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}} \quad (1)$$

di mana $P(Y = 1|X)$ menunjukkan probabilitas mahasiswa merasa puas berdasarkan variabel prediktor X , β_0 adalah intercept, dan β_i adalah koefisien untuk setiap variabel prediktor. Parameter model diestimasi menggunakan *Maximum Likelihood Estimation* (MLE). Regresi logistik menghasilkan prediksi probabilitas yang dapat dikonversi menjadi label kelas (puas atau tidak puas) berdasarkan nilai ambang tertentu, biasanya 0.5 (Menard, 2010).

2.2 Support Vector Machine (SVM)

SVM merupakan algoritma klasifikasi yang bertujuan menemukan hyperplane optimal yang memaksimalkan margin antara dua kelas (Ma dan Guo, 2014). Fungsi keputusan didefinisikan sebagai:

$$f(z) = \text{sign} \left(\sum_{i=1}^p \alpha_i y_i K(x_i, z) + b \right) \quad (2)$$

di mana x_i adalah *support vectors*, y_i adalah label kelas, α_i adalah *Lagrange multipliers*, b adalah bias, dan $K(x_i, z)$ adalah fungsi kernel untuk menangani pemisahan nonlinier. Penelitian ini menggunakan *Radial Basis Function* (RBF) kernel karena mampu menangkap pola nonlinier pada data kepuasan mahasiswa (Wang, 2005).

2.3 Naïve Bayes

Naïve Bayes adalah algoritma klasifikasi probabilistik berbasis Teorema Bayes dengan asumsi independensi kondisional antar fitur (Aldi Florensus Gea, A.F., Basir, C. 2025; Sabry, 2023). Aturan klasifikasi dinyatakan sebagai

$$P(C|F_1, F_2, \dots, F_n) = \frac{P(C) \cdot P(F_1, F_2, \dots, F_n|C)}{P(F_1, F_2, \dots, F_n)} \quad (3)$$

di mana C adalah kelas (kategori kepuasan) dan F_1, F_2, \dots, F_n adalah fitur yang diamati. Karena asumsi independensi antar fitur, likelihood disederhanakan menjadi:

$$P(F_1, F_2, \dots, F_n|C) = \prod_{i=1}^n P(F_i|C) \quad (4)$$

Probabilitas posterior untuk setiap kelas dihitung, dan kelas dengan nilai posterior tertinggi dipilih sebagai prediksi. Dalam penelitian ini, Laplace smoothing digunakan untuk menangani masalah frekuensi nol, dan *5-fold cross validation* diterapkan untuk menentukan parameter model yang optimal (Karabatak, 2015).

2.4 Evaluasi Model

Akurasi klasifikasi merupakan ukuran ketepatan klasifikasi yang menunjukkan performansi teknik klasifikasi secara keseluruhan (Amin dan Mahmoud, 2022). Semakin tinggi akurasi klasifikasi berarti performansi teknik klasifikasi juga semakin baik. Umumnya cara mengukur kinerja klasifikasi menggunakan *confusion matrix*. *Confusion matrix* mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya.

Tabel 2. *Confusion Matrix*

	<i>Predicted Positive Class</i>	<i>Predicted Negative Class</i>
<i>Actual Positive Class</i>	<i>TP (True Positive)</i>	<i>FN (False Negative)</i>
<i>Actual Negative Class</i>	<i>FP (False Positive)</i>	<i>TN (True Negative)</i>

Ukuran yang digunakan untuk mengevaluasi kinerja model berdasarkan *confusion matrix* meliputi beberapa metrik, antara lain akurasi (*accuracy*) yang menunjukkan proporsi prediksi yang benar dari seluruh data, presisi (*precision*) yang mengukur seberapa tepat prediksi positif model, sensitivitas (*recall*) yang menunjukkan kemampuan model dalam mendeteksi data positif sebenarnya, dan spesifisitas (*specificity*) yang menggambarkan kemampuan model dalam mengenali data negatif dengan benar.

2.5 Tahapan Analisis Data

Software atau alat bantu yang digunakan untuk pengolahan data ini adalah *R Studio* dengan paket-paket *tidyverse*, *caret*, *e1071*, dan *klaR*, serta *Microsoft Excel 2019* untuk manajemen data awal. Data yang telah diperoleh kemudian dianalisis sebagai berikut:

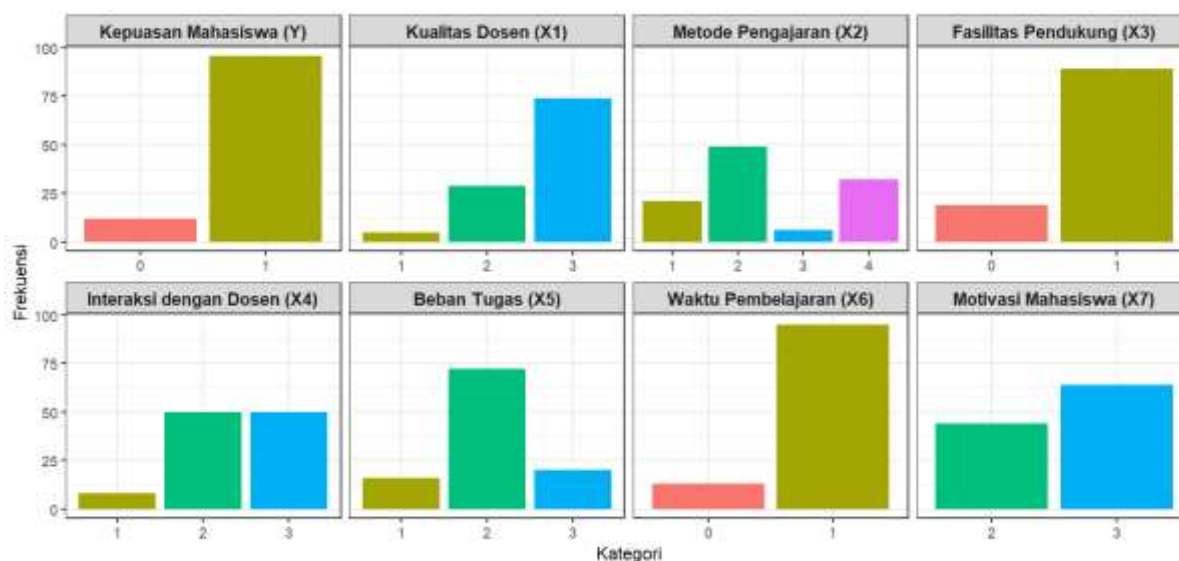
1. **Input Data**
Data mentah hasil survei mahasiswa dimasukkan ke dalam R Studio dan disiapkan untuk analisis.
2. **Pre-processing Data**
Melakukan pembersihan data, mengubah jawaban kuesioner menjadi kategori numerik, serta memastikan semua kolom memiliki format yang sesuai untuk analisis.
3. **Klasifikasi kepuasan mahasiswa menggunakan Regresi Logistik dengan tahapan sebagai berikut:**
 - a. Menentukan variabel prediktor dan variabel respon (kepuasan mahasiswa).
 - b. Membagi dataset menjadi data training (80%) dan data testing (20%) menggunakan *holdout validation*.
 - c. Melatih model regresi logistik pada data training.
 - d. Mengukur performa klasifikasi menggunakan *confusion matrix* dan metrik evaluasi.
4. **Klasifikasi kepuasan mahasiswa menggunakan Support Vector Machine (SVM) dengan tahapan sebagai berikut:**
 - a. Menentukan fungsi kernel (menggunakan RBF) dan nilai parameter kernel.
 - b. Membagi dataset menjadi data training dan data testing dengan *holdout validation*.
 - c. Melakukan *tuning hyperparameter* SVM (misalnya cost dan gamma) menggunakan *cross-validation*.
 - d. Mengukur performa klasifikasi menggunakan *confusion matrix* dan metrik evaluasi.
5. **Klasifikasi kepuasan mahasiswa menggunakan Naïve Bayes dengan tahapan sebagai berikut:**
 - a. Membagi data menjadi training dan testing dengan proporsi 80:20.

- b. Melakukan pelatihan model Naïve Bayes menggunakan *5-fold cross validation*.
 - c. Menerapkan nilai *Laplace smoothing* untuk menangani masalah frekuensi nol.
 - d. Mengukur performa klasifikasi menggunakan confusion matrix dan metrik evaluasi.
6. Perbandingan performa ketiga metode
Membandingkan hasil akurasi dan metrik evaluasi lainnya dari Regresi Logistik, SVM, dan Naïve Bayes pada data testing untuk menentukan algoritma terbaik dalam memprediksi kepuasan mahasiswa.

3. HASIL DAN PEMBAHASAN

3.1 Statistika Deskriptif

Sebelum dilakukan analisis lebih lanjut menggunakan model Logistic Regression, Support Vector Machine (SVM), dan Naïve Bayes, terlebih dahulu dilakukan analisis deskriptif untuk menggambarkan hubungan antara variabel prediktor dan variabel respon, yaitu tingkat kepuasan mahasiswa terhadap proses pembelajaran. Untuk memberikan gambaran yang lebih jelas, berikut disajikan grafik yang menampilkan statistika deskriptif sebagai gambaran umum mengenai data.



Gambar 1. Statistika Deskriptif Variabel Penelitian

Berdasarkan hasil analisis deskriptif, diketahui bahwa mahasiswa yang menyatakan puas terhadap proses pembelajaran memiliki persentase sebesar 88,90%, sedangkan mahasiswa yang tidak puas sebesar 11,10%. Temuan ini menunjukkan bahwa sebagian besar mahasiswa Universitas Papua pada Semester Gasal 2025/2026 menilai proses pembelajaran telah berlangsung dengan baik, meskipun masih terdapat sebagian kecil mahasiswa yang belum merasa puas sepenuhnya.

Ditinjau dari aspek kualitas dosen, mahasiswa yang menilai dosennya berkualitas baik memiliki tingkat kepuasan yang lebih tinggi dibandingkan dengan mahasiswa yang menilai cukup atau kurang baik. Hal ini menunjukkan bahwa kualitas pengajaran dosen merupakan salah satu faktor penting yang berpengaruh terhadap persepsi kepuasan mahasiswa.

Selanjutnya, dari segi metode pengajaran, mahasiswa yang mengikuti pembelajaran dengan pendekatan diskusi dan multimedia cenderung memiliki tingkat kepuasan lebih

tinggi dibandingkan dengan mahasiswa yang mengikuti metode ceramah konvensional. Temuan ini menunjukkan bahwa variasi metode pengajaran yang interaktif dapat meningkatkan keterlibatan mahasiswa serta memperkuat persepsi positif terhadap proses belajar.

Pada aspek fasilitas pendukung pembelajaran, mahasiswa yang menilai fasilitas pembelajaran memadai menunjukkan tingkat kepuasan yang lebih tinggi dibandingkan dengan mahasiswa yang menilai fasilitas tidak memadai. Hal ini mengindikasikan bahwa ketersediaan sarana dan prasarana yang baik masih menjadi faktor penentu utama dalam menciptakan pengalaman belajar yang berkualitas.

Faktor interaksi dengan dosen juga menunjukkan pengaruh yang menarik. Mahasiswa yang menjawab “sering” dan “kadang” berinteraksi dengan dosen memiliki tingkat kepuasan yang relatif sama, yang mengindikasikan bahwa frekuensi interaksi bukan satu-satunya penentu kepuasan. Kemungkinan besar, kualitas interaksi yang terjadi seperti keterbukaan komunikasi dan relevansi umpan balik lebih berperan penting dibandingkan intensitas pertemuan itu sendiri. Dengan demikian, komunikasi yang efektif antara mahasiswa dan dosen memiliki pengaruh yang sebanding dengan frekuensi interaksi.

Dari segi beban tugas, mahasiswa yang menilai beban tugas dalam kategori sedang menunjukkan tingkat kepuasan yang lebih tinggi dibandingkan dengan mereka yang menilai beban tugas terlalu berat. Beban tugas yang proporsional tampaknya mendukung proses pembelajaran yang efektif dan memberikan pengalaman akademik yang lebih seimbang.

Selain itu, mahasiswa yang menilai waktu pelaksanaan pembelajaran sesuai dengan waktu yang dianggap optimal cenderung memiliki tingkat kepuasan lebih tinggi. Hal ini menegaskan bahwa ketepatan jadwal berperan penting dalam menjaga konsistensi kehadiran, fokus belajar, dan kesiapan mahasiswa selama proses pembelajaran.

Terakhir, dari aspek motivasi belajar, mahasiswa dengan tingkat motivasi tinggi menunjukkan kepuasan yang lebih besar dibandingkan dengan mahasiswa yang memiliki motivasi sedang atau rendah. Hal ini memperkuat pandangan bahwa motivasi intrinsik berperan penting dalam membentuk persepsi positif terhadap kualitas pembelajaran yang diterima.

Secara keseluruhan, hasil analisis deskriptif menunjukkan bahwa faktor-faktor seperti kualitas dosen, metode pengajaran, fasilitas pendukung, interaksi dengan dosen, beban tugas, waktu pembelajaran, dan motivasi mahasiswa memiliki hubungan yang searah dengan tingkat kepuasan mahasiswa terhadap proses pembelajaran.

3.2 Evaluasi Model Regresi Logistik

Analisis regresi logistik digunakan dalam penelitian ini untuk memodelkan hubungan antara variabel dependen dan beberapa variabel bebas. Sebelum dilakukan pemodelan, data dibagi menjadi dua bagian menggunakan metode *holdout validation*, yaitu 80% untuk data *training* dan 20% untuk data *testing*. Pembagian ini bertujuan agar model dapat dilatih menggunakan sebagian besar data, sementara sisanya digunakan untuk mengevaluasi kemampuan generalisasi model terhadap data baru.

Model regresi logistik kemudian dibangun menggunakan data *training*. Model ini bertujuan mengidentifikasi pengaruh masing-masing variabel bebas terhadap peluang mahasiswa merasa puas dalam proses pembelajaran. Setelah model terbentuk, dilakukan proses prediksi pada data *training* dan data *testing*. Nilai probabilitas hasil prediksi kemudian dikonversi menjadi kelas kategorik (misalnya: “Puas” atau “Tidak Puas”) menggunakan nilai batas (*cut-off*) 0.5.

Untuk menilai kinerja model, digunakan *confusion matrix*, baik untuk data training maupun

data testing. *Confusion matrix* memberikan gambaran mengenai jumlah prediksi benar dan salah untuk masing-masing kategori. Melalui *confusion matrix* juga diperoleh ukuran performa seperti akurasi, sensitivitas, dan spesifisitas.

Hasil evaluasi ini memungkinkan peneliti untuk memahami seberapa baik model regresi logistik mampu mengklasifikasikan tingkat kepuasan mahasiswa berdasarkan faktor-faktor yang diamati. Dengan demikian, regresi logistik memberikan dasar yang kuat untuk mengidentifikasi variabel yang paling berpengaruh dan menilai ketepatan prediksi model terhadap data baru. Berikut merupakan hasil klasifikasi data *training*.

Tabel 3. Klasifikasi Data *Training* Regresi Logistik

<i>Predicted</i>	<i>Reference</i>		Total
	1	0	
1	77	6	83
0	0	4	4
Total	77	10	87

Dari *confusion* matriks tersebut, diperoleh nilai TN: 4 kasus yang diklasifikasikan dengan benar sebagai tidak puas (kelas 0), FP: 6 kasus yang diprediksi puas (kelas 1) padahal sebenarnya tidak puas (kelas 0), FN: 0 kasus yang diprediksi sebagai tidak puas (kelas 0) padahal sebenarnya puas (kelas 1), TP: 77 kasus yang diklasifikasikan dengan benar sebagai puas (kelas 1). Setelah diperoleh hasil klasifikasi data *training*, berikut merupakan hasil klasifikasi data *testing*.

Tabel 4. Klasifikasi Data Testing Regresi Logistik

Predicted	Reference		Total
	1	0	
1	19	1	20
0	0	1	1
Total	19	2	21

Dari *confusion* matriks tersebut, diperoleh nilai TN: 1 kasus yang diklasifikasikan dengan benar sebagai tidak puas (kelas 0), FP: 1 kasus yang diprediksi puas (kelas 1) padahal sebenarnya tidak puas (kelas 0), FN: 0 kasus yang diprediksi sebagai tidak puas (kelas 0) padahal sebenarnya puas (kelas 1), TP: 19 kasus yang diklasifikasikan dengan benar sebagai puas (kelas 1). Selanjutnya diperoleh matriks kinerja model sebagai berikut.

Tabel 5. Kinerja Model

Metriks	Data <i>Training</i>	Data <i>Testing</i>
<i>Accuracy</i>	0.9310	0.9524
<i>Precision</i>	0.9277	0.9500
<i>Sensitivity</i>	1.0000	1.0000
<i>Specificity</i>	0.4000	0.5000

Berdasarkan hasil pada Tabel 5, dengan nilai *accuracy* data *training* adalah 93.10% menunjukkan bahwa model memiliki performa yang baik dalam mengklasifikasikan data secara keseluruhan, *precision* sebesar 92.77% menunjukkan dari semua prediksi positif adalah benar, *sensitivity* sebesar 1.0000 berarti model berhasil mengidentifikasi 100% dari

semua kelas positif (puas) yang sebenarnya, dan *specificity* sebesar 0.4000 berarti model berhasil mengidentifikasi 40.00% (tidak puas) dari semua kelas negatif yang sebenarnya.

Nilai *accuracy* data *testing* adalah 95.24% menunjukkan bahwa model memiliki performa yang baik dalam mengklasifikasikan data secara keseluruhan, *precision* sebesar 95.00% menunjukkan dari semua prediksi positif adalah benar, *sensitivity* sebesar 1.000 berarti model berhasil mengidentifikasi 100% dari semua kelas positif (puas) yang sebenarnya, dan *specificity* sebesar 0.5000 berarti model berhasil mengidentifikasi 50.00% (tidak puas) dari semua kelas negatif yang sebenarnya. Secara keseluruhan, hasil evaluasi ini menunjukkan bahwa model memiliki performa yang sangat baik dalam memprediksi kepuasan mahasiswa, dengan tingkat akurasi, presisi, dan sensitivitas yang tinggi.

3.3 Evaluasi Model Support Vector Machine

Pada analisis Support Vector Machine (SVM) dengan kernel *Radial Basis Function* (RBF), proses pembangunan model dilakukan dengan menentukan hyperparameter berupa nilai C (*cost*) dan γ (*gamma*). Kedua parameter tersebut memiliki peran penting dalam mengendalikan kompleksitas model, margin pemisah, serta sensitivitas model terhadap data.

Untuk memperoleh kombinasi parameter terbaik, dilakukan proses *tuning hyperparameter* secara simultan menggunakan pendekatan *grid search*, yaitu mencoba seluruh kombinasi nilai C dan γ yang telah ditetapkan sebelumnya. Pada penelitian ini, nilai parameter yang diuji terdiri dari beberapa kandidat, sehingga menghasilkan sejumlah kombinasi hyperparameter yang memungkinkan.

Proses *tuning hyperparameter* dilaksanakan menggunakan metode *5-fold cross validation*, sehingga setiap kombinasi parameter dievaluasi sebanyak lima kali untuk mendapatkan nilai akurasi yang stabil dan tidak bias. Kombinasi parameter yang memberikan akurasi validasi tertinggi dipilih sebagai model terbaik. Berdasarkan hasil *tuning* tersebut, diperoleh bahwa kombinasi parameter C dan γ tertentu (misalnya hasil berjalan Anda: $C = 1$ dan $\gamma = 0.01$) merupakan kombinasi yang menghasilkan performa paling optimal berdasarkan akurasi validasi.

Model terbaik tersebut kemudian digunakan untuk membangun model SVM final. Selanjutnya, dataset dibagi menggunakan metode *holdout validation* dengan proporsi 70% data *training* dan 30% data *testing*. Data *training* digunakan untuk melatih model SVM terbaik, sedangkan data *testing* digunakan untuk mengevaluasi performa model dalam melakukan prediksi pada data baru.

Setelah model dilatih, dilakukan proses prediksi untuk kedua subset data (*training* dan *testing*), kemudian dievaluasi menggunakan *confusion matrix*. *Confusion matrix* tersebut memberikan informasi terkait jumlah prediksi benar dan salah pada masing-masing kelas sehingga dapat menilai akurasi model, sensitivitas, spesifisitas, dan kinerja klasifikasi secara keseluruhan.

Dengan demikian, tahapan analisis SVM telah menghasilkan model terbaik berdasarkan kombinasi hyperparameter optimal dan evaluasi menggunakan data *training* maupun *testing*. Berikut merupakan hasil klasifikasi data *training*.

Tabel 6. Klasifikasi Data Training SVM

Predicted	Reference		
	1	0	Total
1	77	6	83
0	0	4	4
Total	77	10	87

Dari *confusion matrix* tersebut, diperoleh nilai TN: 4 kasus yang diklasifikasikan dengan benar sebagai tidak puas (kelas 0), FP: 6 kasus yang diprediksi puas (kelas 1) padahal sebenarnya tidak puas (kelas 0), FN: 0 kasus yang diprediksi sebagai tidak puas (kelas 0) padahal sebenarnya puas (kelas 1), TP: 77 kasus yang diklasifikasikan dengan benar sebagai puas (kelas 1). Setelah diperoleh hasil klasifikasi data *training*, berikut merupakan hasil klasifikasi data *testing*.

Tabel 7. Klasifikasi Data Testing SVM

<i>Predicted</i>	<i>Reference</i>		Total
	1	0	
1	19	1	20
0	0	1	1
Total	19	2	21

Dari *confusion matrix* tersebut, diperoleh nilai TN: 1 kasus yang diklasifikasikan dengan benar sebagai tidak puas (kelas 0), FP: 1 kasus yang diprediksi puas (kelas 1) padahal sebenarnya tidak puas (kelas 0), FN: 0 kasus yang diprediksi sebagai tidak puas (kelas 0) padahal sebenarnya puas (kelas 1), TP: 19 kasus yang diklasifikasikan dengan benar sebagai puas (kelas 1). Selanjutnya diperoleh matriks kinerja model sebagai berikut.

Tabel 8. Kinerja Model Klasifikasi SVM

Metriks	Data <i>Training</i>	Data <i>Testing</i>
Accuracy	0.9310	0.9524
Precision	0.9277	0.9500
Sensitivity	1.000	1.000
Specificity	0.400	0.500

Berdasarkan hasil pada Tabel 8, dengan nilai *accuracy* data *training* adalah 93.10% menunjukkan bahwa model memiliki performa yang baik dalam mengklasifikasikan data secara keseluruhan, *precision* sebesar 92.77% menunjukkan dari semua prediksi positif adalah benar, *sensitivity* sebesar 1.0000 berarti model berhasil mengidentifikasi 100% dari semua kelas positif (puas) yang sebenarnya, dan *specificity* sebesar 0.4000 berarti model berhasil mengidentifikasi 40.00% (tidak puas) dari semua kelas negatif yang sebenarnya.

Nilai *accuracy* data *testing* adalah 95.24% menunjukkan bahwa model memiliki performa yang baik dalam mengklasifikasikan data secara keseluruhan, *precision* sebesar 95.00% menunjukkan dari semua prediksi positif adalah benar, *sensitivity* sebesar 1.000 berarti model berhasil mengidentifikasi 100% dari semua kelas positif (puas) yang sebenarnya, dan *specificity* sebesar 0.5000 berarti model berhasil mengidentifikasi 50.00% (tidak puas) dari semua kelas negatif yang sebenarnya. Secara keseluruhan, hasil evaluasi ini menunjukkan bahwa model memiliki performa yang sangat baik dalam memprediksi kepuasan mahasiswa, dengan tingkat akurasi, presisi, dan sensitivitas yang tinggi.

3.4 Evaluasi Model Naïve Bayes

Metode Naïve Bayes merupakan algoritma klasifikasi berbasis probabilistik yang menggunakan Teorema Bayes sebagai dasar perhitungan dalam menentukan kelas suatu observasi. Model ini mengasumsikan bahwa seluruh variabel prediktor bersifat independen satu sama lain dengan variabel target. Walaupun asumsi independensi tersebut tidak selalu terpenuhi secara sempurna dalam data empiris, Naïve Bayes tetap dikenal efektif, sederhana,

serta mampu bekerja baik meskipun jumlah data tidak terlalu besar.

Dalam penelitian ini, metode Naïve Bayes digunakan untuk memprediksi kepuasan mahasiswa terhadap proses pembelajaran dengan melibatkan berbagai variabel prediktor seperti kualitas dosen, metode pengajaran, fasilitas pendukung, interaksi dengan dosen, beban tugas, kesesuaian waktu pembelajaran, dan motivasi mahasiswa. Model ini dipilih karena kemampuannya dalam menghitung probabilitas dari masing-masing kategori secara langsung dan interpretasinya yang ringkas.

Untuk membangun model yang lebih stabil, terutama ketika terdapat kategori jawaban tertentu yang jarang muncul, digunakan teknik *Laplace smoothing*. Pada penelitian ini, nilai *Laplace* = 1 ditetapkan sebagai parameter terbaik berdasarkan hasil *Cross Validation*. Teknik *smoothing* ini mencegah probabilitas menjadi nol pada kategori yang frekuensinya sangat kecil sehingga model tetap dapat melakukan prediksi secara konsisten.

Proses pemilihan parameter terbaik dilakukan menggunakan metode *5-Fold Cross Validation* (CV). Data dibagi menjadi 5 bagian yang kurang lebih sama besar. Selanjutnya, model dilatih pada 4 bagian data dan diuji pada 1 bagian yang tersisa. Proses ini diulang sebanyak 5 kali hingga setiap bagian digunakan sebagai data uji sekali.

Pendekatan ini memberikan estimasi performa model yang lebih stabil serta membantu menghindari *overfitting*. Dari hasil CV, nilai *Laplace* terbaik dipilih berdasarkan nilai akurasi rata-rata tertinggi dari kelima iterasi tersebut.

Setelah parameter ditentukan, model Naïve Bayes dibangun menggunakan data *training* (80%) dan kemudian diuji pada data testing (20%). Evaluasi performa dilakukan menggunakan *confusion matrix*, yang menggambarkan tingkat ketepatan model dalam memprediksi kategori kepuasan mahasiswa. Melalui *confusion matrix*, dapat dihitung berbagai metrik seperti akurasi, sensitivitas, dan spesifisitas. Hasil ini kemudian dibandingkan dengan performa model Regresi Logistik dan Support Vector Machine (SVM) untuk mengetahui metode yang paling optimal dalam memodelkan kepuasan mahasiswa. Berikut merupakan hasil klasifikasi data *training*.

Tabel 9. Klasifikasi Data Training Naïve Bayes

Predicted	Reference		Total
	1	0	
1	66	5	71
0	1	3	4
Total	67	8	75

Dari *confusion matrix* tersebut, diperoleh nilai TN: 3 kasus yang diklasifikasikan dengan benar sebagai tidak puas (kelas 0), FP: 5 kasus yang diprediksi puas (kelas 1) padahal sebenarnya tidak puas (kelas 0), FN: 1 kasus yang diprediksi sebagai tidak puas (kelas 0) padahal sebenarnya puas (kelas 1), TP: 66 kasus yang diklasifikasikan dengan benar sebagai puas (kelas 1). Setelah diperoleh hasil klasifikasi data *training*, berikut merupakan hasil klasifikasi data *testing*.

Tabel 10. Klasifikasi Data Testing Naïve Bayes

Predicted	Reference		Total
	1	0	
1	27	2	29
0	2	2	4
Total	29	4	31

Dari *confusion matrix* tersebut, diperoleh nilai TN: 2 kasus yang diklasifikasikan dengan benar sebagai tidak puas (kelas 0), FP: 2 kasus yang diprediksi puas (kelas 1) padahal sebenarnya tidak puas (kelas 0), FN: 2 kasus yang diprediksi sebagai tidak puas (kelas 0) padahal sebenarnya puas (kelas 1), TP: 27 kasus yang diklasifikasikan dengan benar sebagai puas (kelas 1). Selanjutnya diperoleh matriks kinerja model sebagai berikut.

Tabel 11. Kinerja Model Klasifikasi Naïve Bayes

Metriks	Data <i>Training</i>	Data <i>Testing</i>
<i>Accuracy</i>	0.9200	0.8788
<i>Precision</i>	0.9296	0.9310
<i>Sensitivity</i>	0.9851	0.9310
<i>Specificity</i>	0.3750	0.5000

Berdasarkan hasil pada Tabel 11, dengan nilai *accuracy* data *training* adalah 92.00% menunjukkan bahwa model memiliki performa yang baik dalam mengklasifikasikan data secara keseluruhan, *precision* sebesar 92.96% menunjukkan dari semua prediksi positif adalah benar, *sensitivity* sebesar 0.9851 berarti model berhasil mengidentifikasi 98.51% dari semua kelas positif (puas) yang sebenarnya, dan *specificity* sebesar 0.3750 berarti model berhasil mengidentifikasi 37.50% (tidak puas) dari semua kelas negatif yang sebenarnya.

Nilai *accuracy* data *testing* adalah 87.88% menunjukkan bahwa model memiliki performa yang baik dalam mengklasifikasikan data secara keseluruhan, *precision* sebesar 93.10% menunjukkan dari semua prediksi positif adalah benar, *sensitivity* sebesar 0.9310 berarti model berhasil mengidentifikasi 93.10% dari semua kelas positif (puas) yang sebenarnya, dan *specificity* sebesar 0.5000 berarti model berhasil mengidentifikasi 50.00% (tidak puas) dari semua kelas negatif yang sebenarnya. Secara keseluruhan, hasil evaluasi ini menunjukkan bahwa model memiliki performa yang baik dalam memprediksi kepuasan mahasiswa, dengan tingkat akurasi, presisi, dan sensitivitas yang tinggi.

4. KESIMPULAN

Metode Regresi Logistik dan SVM menunjukkan performa yang sama, dengan nilai akurasi sebesar 95.24%, yang mengindikasikan bahwa kedua metode mampu mengklasifikasikan kepuasan mahasiswa dengan sangat baik. SVM tetap unggul dari sisi fleksibilitas model dalam menangkap pola data yang lebih kompleks ketika dilakukan tuning hyperparameter. Sementara itu, metode Naïve Bayes, yang menggunakan Laplace smoothing dan *5-fold cross validation* dalam penentuan parameter terbaik, menghasilkan akurasi sebesar 87.88%. Meskipun akurasinya lebih rendah dibandingkan dua metode lainnya, Naïve Bayes tetap menawarkan keunggulan dalam hal kesederhanaan, efisiensi komputasi, dan stabilitas prediksi.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa baik Regresi Logistik maupun SVM merupakan metode yang paling efektif untuk memprediksi kepuasan mahasiswa pada dataset ini, dengan performa yang setara. Temuan ini dapat menjadi dasar bagi institusi pendidikan dalam memilih pendekatan analisis yang sesuai, baik untuk tujuan interpretasi model (Regresi Logistik) maupun akurasi prediksi yang optimal (SVM). Sementara itu, Naïve Bayes tetap relevan digunakan pada kondisi yang membutuhkan pemodelan cepat dan sederhana.

5. DAFTAR PUSTAKA

- Seila Amalia, Banjarnahor, R. M., Sinaga, M. R. T., Sianturi, A. A., & Dinie Triana. (2025). Model Regresi Logistik Pengaruh Aksesibilitas dan Kondisi Sosial-Ekonomi Terhadap Ketimpangan Pendidikan di Wilayah Desa Bagan Serdang Kec. Pantai Labu. *STATMAT: JURNAL STATISTIKA DAN MATEMATIKA*, 7(2), 170–179. Retrieved from <https://openjournal.unpam.ac.id/index.php/sm/article/view/52010>
- Amin F, Mahmoud M. Confusion matrix in binary classification problems: A step-by-step tutorial. *J Eng Res*. 2022;6(5).
- Aristawidya R, Indahwati I, Erfiani E, Fitrianto A, AA M. Perbandingan analisis regresi logistik biner dan naïve bayes classifier untuk memprediksi faktor risiko diabetes. *J Lebesgue*. 2024;5(2):782–94.
- Cahyani R, Putri SA, Lestari D. Prediksi risiko penyakit diabetes menggunakan algoritma regresi logistik. *J Teknol Inf dan Komput*. 2022;7(2):115–24.
- Fahrezi RA, Wijaya MY, Fitriyati N. Prediksi harga penutupan saham Bank Central Asia: implementasi algoritma long short-term memory dan perbandingannya dengan support vector machine. *J Lebesgue*. 2024;5(1):452–64.
- Aldi Florensus Gea, A.F., Basir, C. 2025. Implementasi Naïve Bayes Classifier (NBC) untuk Menganalisis Sentimen Ulasan Film “Everything Everywhere All at Once” di Twitter. *J. Ris. & Ap. Mat*. Vol. 9 No. 2 (2025) pp. 208-218. <https://doi.org/10.26740/jram.v9n2.p208-218>
- Hosmer DW Jr, Lemeshow S, Sturdivant RX. Applied logistic regression. New York: John Wiley & Sons; 2013.
- Karabatak M. A new classifier for breast cancer detection based on Naïve Bayesian. *Measurement*. 2015;72:32–6.
- Ma Y, Guo G. Support vector machines applications. Berlin: Springer; 2014.
- Menard SW. Logistic regression: from introductory to advanced concepts and applications. Thousand Oaks: SAGE; 2010.
- Roihan A, Sunarya PA, Rafika AS. Pemanfaatan machine learning dalam berbagai bidang. *J Khatulistiwa Inform*. 2020;5(1):1–10.
- Roni S, Atmaja T. Analisis faktor-faktor yang mempengaruhi kepuasan mahasiswa di Universitas Potensi Utama. *J Menara Ekon*. 2024;10(2).
- Sabry F. Naive Bayes classifier: fundamentals and applications. One Billion Knowledgeable; 2023.
- Sari FR, Fitri F, Putra AA, Permana D. Comparison of Naive Bayes method and binary logistic regression on classification of social assistance recipients Program Keluarga Harapan (PKH). *UNP J Stat Data Sci*. 2023;1(2):82–9.
- Wang L. Support vector machines: theory and applications. Berlin: Springer; 2005.